

A6_RegresionPoisson

Josue Salvador Cano Martinez

2022-11-06

Cargar datos

```
# Base de datos warpbreaks que contiene datos del hilo para identificar cuáles variables predictoras af  
data<-warpbreaks  
head(data,10)
```

```
##      breaks wool tension  
## 1         26    A      L  
## 2         30    A      L  
## 3         54    A      L  
## 4         25    A      L  
## 5         70    A      L  
## 6         52    A      L  
## 7         51    A      L  
## 8         26    A      L  
## 9         67    A      L  
## 10        18    A      M
```

Análisis de base de datos

Descripción de variables y número de datos. Descripción de los valores que toma y que tipo de variable son

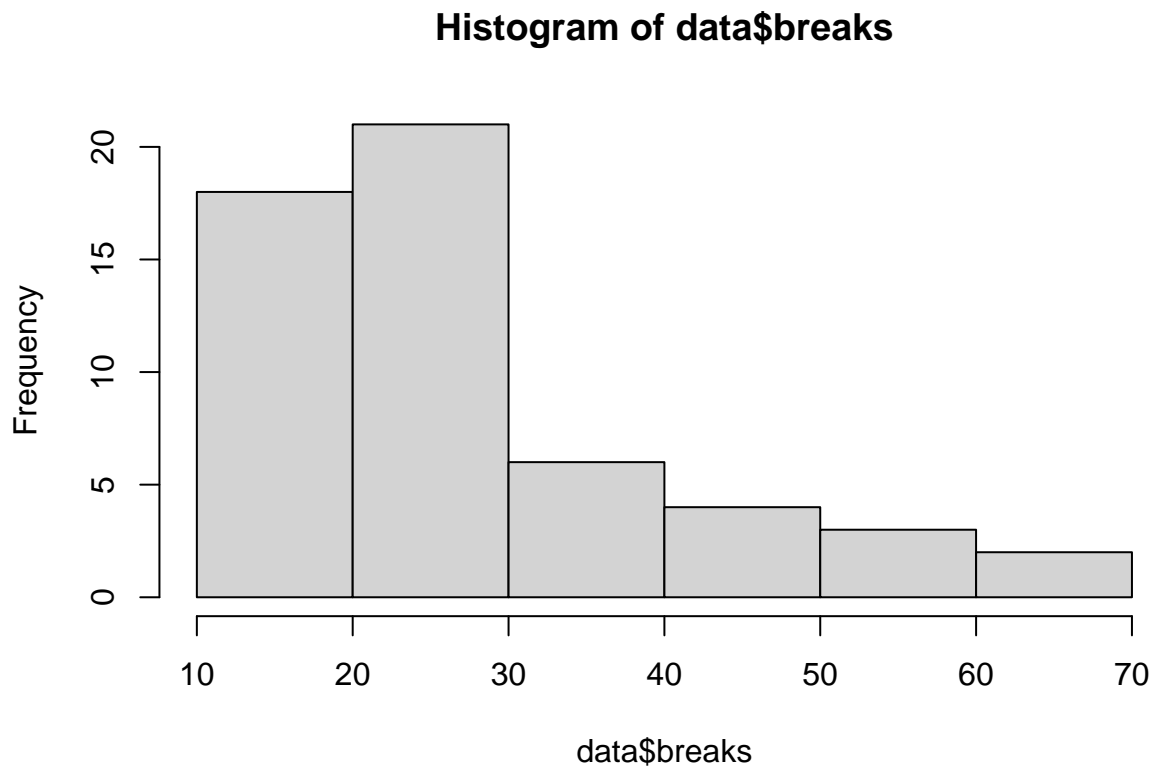
```
ls.str(data)
```

```
## breaks :  num [1:54] 26 30 54 25 70 52 51 26 67 18 ...  
## tension :  Factor w/ 3 levels "L","M","H": 1 1 1 1 1 1 1 1 1 2 ...  
## wool :    Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ...
```

```
# breaks es la variable respuesta que representa el número de rupturas, mientras que wool (tipo de lana.
```

Análisis del histograma del número de rupturas

```
hist(data$breaks)
```



Se puede observar que la distribución de los datos no sigue una distribución normal.

Media y varianza del número de rupturas

```
# Calcular media  
mean(data$breaks)
```

```
## [1] 28.14815
```

```
# Calcular varianza  
var(data$breaks)
```

```
## [1] 174.2041
```

La media y la varianza son diferentes, se puede apreciar que la varianza es mucho mayor que la media,

Ajuste del modelo de regresión Poisson

```
poisson.model<-glm(breaks ~ wool + tension, data, family = poisson(link = "log"))
summary(poisson.model)
```

```
##
## Call:
## glm(formula = breaks ~ wool + tension, family = poisson(link = "log"),
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.69196    0.04541  81.302  < 2e-16 ***
## woolB        -0.20599    0.05157  -3.994  6.49e-05 ***
## tensionM     -0.32132    0.06027  -5.332  9.73e-08 ***
## tensionH     -0.51849    0.06396  -8.107  5.21e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 210.39  on 50  degrees of freedom
## AIC: 493.06
##
## Number of Fisher Scoring iterations: 4
```

Como se puede apreciar, el p-value es menor que 0.05, lo cual permite entender el impacto de las vari

Debido a que la desviación residual es mayor que los grados de libertad, se puede entender que existe

```
poisson.model2<-glm(breaks ~ wool + tension, data = data, family = quasipoisson(link = "log"))
summary(poisson.model2)
```

```
##
## Call:
## glm(formula = breaks ~ wool + tension, family = quasipoisson(link = "log"),
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.69196    0.09374  39.384  < 2e-16 ***
## woolB        -0.20599    0.10646  -1.935  0.058673 .
##
```

```
## tensionM      -0.32132      0.12441     -2.583 0.012775 *
## tensionH      -0.51849      0.13203     -3.927 0.000264 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.261537)
##
##      Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 210.39  on 50  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Para tener un error estándar más correcto se hace uso de un modelo quasipoisson.

Comparación de modelos

```
library(arm)
```

```
## Warning: package 'arm' was built under R version 4.2.2
```

```
## Loading required package: MASS
```

```
## Loading required package: Matrix
```

```
## Loading required package: lme4
```

```
##
```

```
## arm (Version 1.13-1, built: 2022-8-25)
```

```
## Working directory is C:/Users/josue/Documents/OneDrive - Instituto Tecnológico y de Estudios Superiores de
```

```
# Coeficientes del primer modelo
```

```
coefFirstModel = coef(poisson.model)
```

```
# Coeficientes del segundo modelo
```

```
coefSecondModel = coef(poisson.model2)
```

```
# Errores estándar del primer modelo
```

```
errFirstModel = se.coef(poisson.model)
```

```
# Errores estándar del segundo modelo
```

```
errSecondModel = se.coef(poisson.model2)
```

```
# Combinación de valores
```

```
twoModels <- cbind(coefFirstModel, errFirstModel, coefSecondModel, errSecondModel, exponent = exp(coefF
twoModels
```

```
##               coefFirstModel errFirstModel coefSecondModel errSecondModel
## (Intercept)      3.6919631      0.04541069      3.6919631      0.09374352
## woolB            -0.2059884      0.05157117     -0.2059884      0.10646089
## tensionM         -0.3213204      0.06026580     -0.3213204      0.12440965
## tensionH         -0.5184885      0.06395944     -0.5184885      0.13203462
```

```
##          exponent
## (Intercept) 40.1235380
## woolB      0.8138425
## tensionM   0.7251908
## tensionH   0.5954198
```

En el resultado anterior se puede apreciar que los coeficientes son los mismos, pero los errores estándar son diferentes. Teniendo en cuenta esto, la estimación de la lana tiene un valor de '0.2059884 con un exponente de 0.8138425. Lo anterior nos dice que el cambio de tipo de lana A a tipo de lana B da como resultado una disminución de roturas 0.8138425 veces la intersección porque la estimación del coeficiente es negativa, es decir, si se cambia el tipo de lana A a B, el número de roturas caerá un 18% suponiendo que las demás variables permanecen iguales.